

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 15:52:19 ; Search time 170.72 Seconds
(without alignments)
27.443 Million cell updates/sec

Title: US-09-331-631A-1_COPY_117_185

Perfect score: 384
Sequence: 1 NRORPDQOQYEQCCOKHCORR.....EEQOREDEKYEERMKERDN 69

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Query length	DB ID	Description
1	172	44.8	566	2 S22477	vicillin precursor
2	152.5	39.7	588	1 FMCNAB	alpha-globulin B p
3	147	38.3	509	2 S08059	alpha-globulin typ
4	130.5	34.0	605	2 S06398	alpha-globulin typ
5	116.5	30.3	1038	2 T02634	rep protein homolo
6	114	29.7	810	2 T44430	protein PV100 (imp
7	105	27.3	613	2 S27770	hypothetical prote
8	105	27.3	1737	2 A59235	unconventional myo
9	104.5	27.2	385	2 T19201	hypothetical prote
10	104	27.1	1390	2 T14004	hypothetical prote
11	103.5	27.0	1898	1 A45973	trichohyalin - hum
12	102	26.6	1407	1 S28589	trichohyalin - hum
13	101	26.3	233	2 T17218	hypothetical prote
14	101	26.3	1027	2 T46481	hypothetical prote
15	101	26.3	1233	2 T30989	hypothetical prote
16	98	25.5	648	1 J01150	serine/threonine p
17	97.5	25.4	839	2 T30989	protein kinase (EC
18	97.5	25.4	877	2 I50590	class I INCEMP pr
19	95	24.7	385	2 T20410	hypothetical prote
20	95	24.7	524	2 J01730	62k sucrose-bindin
21	95	24.7	781	2 T02272	hypothetical prote
22	95	24.7	849	2 S61962	probable membrane
23	95	24.7	1023	2 S12519	glutactin - fruit
24	95	24.7	1085	2 S62516	glutactin - fruit
25	94	24.5	285	1 I46207	hypothetical colle
26	94	24.5	678	2 A54514	involucrin - dog
27	93	24.2	529	2 A71899	glutamic acid-rich
28	92.5	24.1	538	2 S29521	hypothetical prote
29	92	24.0	338	2 S04321	casein kinase I ho
					legumin B (clone p

30	92	24.0	429	2 S29565	apolipoprotein A-I
31	92	24.0	523	2 T24961	hypothetical prote
32	92	24.0	566	2 T06453	probable legumin B
33	92	24.0	905	1 RGHYS5	regulatory protein
34	92	24.0	1403	2 S24548	homocytic protein p
35	91.5	23.8	406	2 T24492	hypothetical prote
36	91.5	23.8	550	2 A46419	trophoblast-endoth
37	91.5	23.8	695	2 I54325	gene xzf protein -
38	91	23.7	139	2 A26892	Mopa box protein -
39	91	23.7	483	2 T06459	62k sucrose-bindin
40	91	23.7	1344	2 T42637	hypothetical prote
41	90.5	23.6	758	2 S54522	hypothetical prote
42	90.5	23.6	1263	2 T15496	hypothetical prote
43	90	23.4	1549	1 A40691	trichohyalin - she
44	90	23.4	1905	2 T18267	multidrug resistan
45	89.5	23.3	393	2 J06179	dorsal switch prot

ALIGNMENTS

RESULT 1
S22477
vicillin precursor - cacao
C:Species: Theobroma cacao (cacao)
C:Date: 30-Sep-1993 #sequence-revision 30-Sep-1993 #text-change 17-Mar-1999
C:Accession: S22477; S22478; S18105; S22050
R:McHenry, L., Fritz, P.J., 1173-1176, 1992
Plant Mol. Biol. 18, 1173-1176, 1992
A:Title: Comparison of the structure and nucleotide sequence of vicillin genes of coco
A:Reference number: S22477; UID:92288309
A:Accession: S22477
A:Molecule type: DNA
A:Residues: 1-566 <MCU>
A:Cross-references: EMBL:X62625
A:Accession: S22478
A:Molecule type: mRNA
A:Residues: 1-452 <MC2>
A:Cross-references: EMBL:X62626
C:Genetics:
A:introns: 211/1; 269/3; 296/3; 391/3; 502/1
C:Superfamily: glycinin
C:Keywords: seed; storage protein
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-566/Product: vicillin #status predicted <MAT>

Query Match 44.8%; Score 172; DB 2; Length 566;
Best Local Similarity 31.4%; Pred. No. 2.6e-07;
Matches 32; Conservative 20; Mismatches 14; Indels 36; Gaps 2;

OY 3 ORPDQOQYEQCCOKHCORRTERPRHMTCCORCERREKRRKQO-----46
DB 35 ERPDQOQYEQCCORCERSEATEREERQCCORCERREKRRKQO-----46
OY 47 -----KRYEEOREDEKYE-----EERMKERD 68
DB 95 QQQGQREDEKYECCORCERSEATEREERQCCORCERREKRRKQO-----136

RESULT 2
FMCNAB
alpha-globulin B precursor (clone C72) - upland cotton
A:Alternate names: seed storage protein; vicillin precursor
C:Species: Gossypium hirsutum (upland cotton)
C:Date: 30-Sep-1991 #sequence-revision 30-Sep-1991 #text-change 16-Jul-1999
C:Accession: A30838; S06911
R:Chlan, C.A.; Pyle, J.B.; Legocki, A.B.; Dure III, L.
Plant Mol. Biol. 7, 475-489, 1986
A:Title: Developmental biochemistry of cottonseed embryogenesis and germination XVIII
A:Reference number: A30838
A:Accession: A30838
A:Molecule type: mRNA

RESULT 11

A:Accession: A45973

trichohyalin - human

C:Species: Homo sapiens (man)

C:Date: 03-May-1994 #sequence_revision 01-Mar-1996 #text_change 22-Jun-1999

C:Accession: A45973

R:Lee, S.C.; Kim, I.G.; Markov, L.N.; O'Keefe, E.J.; Parry, D.A.D.; Steinert, P.M.

J. Biol. Chem. 268, 12164-12176, 1993

A:Title: The structure of human trichohyalin. Potential multiple roles as a functional head (cross-linking) protein.

A:Reference number: A45973; MUID:93280194

A:Accession: A45973

A:Molecule type: DNA

A:Residues: 1-1898 <LEE>

A:Cross-references: GB:I09190; NID:g292835; PIDN:AAA65582.1; PID:g292836

A:Note: authors translated the codon AGG for residue 1714 as Pro

C:Comment: Trichohyalin is a protein of the medulla of the hair and of the inner root sheath. Covalent modifications to this protein include conversion of arginine to citrulline and C:Genetics:

A:Gene: GDB:TMH

A:Cross-references: GDB:I36223; OMIM:190370

A:Map position: 1q21-1q21

C:Superfamily: trichohyalin repeat homology

C:Keywords: calcium binding; calmodulin; caltulline; EF hand; hair; tandem repeat

F:49-81/Domain: calmodulin repeat homology <EF2>

Query Match	27.0%;	Score 103.5;	DB 1;	Length 1898;
Best Local Similarity	36.2%;	Pred. No. 0.26;		
Matches 25; Conservative	19;	Mismatches 18;	Indels 7;	Gaps 3

```

0y      2  RQVRDQDQYE-QSCKNCRKREPRNHOCSQDQREKRYEKKRQDQKRYEED--RDEE 58
      || : ||| : : : : || : : || : : || : ||| ||| |||
Db     290 RQEEQDQDQLRQEQDLRKQDEEER---EQDEERQEQDERQEQDERRQEQDLRQDEE 345

```

```
QY      59 KYEERMKEE 67
          : |::: |
Db      346 RREQLRRE 354
```

RESULT 12
S28589
trichohyalin - rabbit
C:Species: *Oryctolagus cuniculus* (domestic rabbit)
C:Date: 12-Mar-1993 #sequence_revision 01-Mar-1996 #text_change 22-Jun-1996
C:Accession: S28589
R:Flatz, M.J.; Rogers, G.E.
Submitted to the EMBL Data Library, December 1992
;Description: Examination of the gene encoding rabbit trichohyalin.

A:Reference number: 52603
 A:Accession: S28589
 A:Molecule type: DNA
 A:Residues: 1-1407 <FE>
 A:Cross-references: EMBL:Z19092; NID:g1746; PIND:CAA79519.1; PID:g1747
 C:Comment: Trichohyalin is a protein of the medulla of the hair and of the inner root sheath.
 C:Covalent modifications to this protein include conversion of arginine to citrulline and
 C:Genetics:
 A:Introns: 46/3
 C:Superfamily: trichohyalin; calmodulin repeat homology
 C:Keywords: calcium binding; citrulline; EF hand; hair; tandem repeat
 E:43-81/Domain: calmodulin repeat homology <EF2>

Query Match	26.6%	Score 102;	DB 1;	Length 1407;
Best Local Similarity	33.8%	Pred. No. 0.28;		
Matches	25;	Conservative	20;	Mismatches 19;
				Indels 10;
				Gaps 2;

```

0Y      3 QRDPOOYEECCOKHCORRETERPHMOTCQQRCERRYKKEKKRQOK-----RYEEQR 54
      | : : : : | | | | | | | | | | : : : : | | :
Db 256 QQQLRLEIRERERQRLQEGERREQ--QLRRQRLQEGERRERQQLRLEIRERERQL 313

```

QY 55 EDEEKYEERMKEED 68

```
Db      314 EQERREQRLEQE 327
```

RESULT 13
T117218
hypothetical protein DKFZp434p1750.1 - human
C:Species: Homo sapiens (man)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T117218
R:Postka, A., Klein, M., Mewes, H.W., Gassenhuber, J., Wiemann, S.
submitted to the Protein Sequence Database, September 1999
A:Reference number: Z18723
A:Accession: T117218
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-233 <POU>
A:Cross-references: EMBL:AL117408
A:Experimental source: adult testis; clone DKFZp434p1750
C:Genetics:
A:Note: DKFZp434p1750.1

Query Match	26.3%;	Score 101;	DB 2;	Length 233;
Best Local Similarity	35.8%;	Pred. No. 0.08;		
Matches: 19;	Conservative 18;	Mismatches 16;	Indels 0;	Gaps 0;

```
Oy      15 KHCQRRETEPRHMOTCOQRCBRRYEKEKROOKRYEEOBREDEEKVEERMKEE 67
          |::|:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db     144 KEMRQKEKERQKQEKERQKQEKEREKEREKQKQEKEREKQEKEREKQKQEF 196
```

RESULT 14
T46481
hypothetical protein DKFzp434A025.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C:Accession: T46481
R:Diesterhoef, A.; Lauber, J.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, January 2000
A:Reference number: Z23035
A:Accession: T46481
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1027 <AA>
A:Cross-references: EMBL:AL137755
A:Experimental source: adult testis; clone DKFzp434A025
C:Genetics:
A:Note: DKFzp434A025.1

Query Match 26.3%; Score 101; DB 2; Length 1027;
Best Local Similarity 36.1%; Pred. No. 0.26;
Matches 30; Conservative 16; Mismatches 19; Indels 18; Gaps 4;

```

0y 2 RQRPQOQOYQCSQK--HCQRRETPRHQOTCOQHCERYEKEKRR---OQKRYEEOQ- 54
   ||: ||| ||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 134 RQKREQOQKQCRRLLEQQRREARQOQERQ--RRQEERKRLLELERRRKEEERR 191

```

```
QY      55 -----EDEEKYEERMKEED 68
          | : : | | :
Db      192 RAE EKRRVEREQEYIRRLQLEEE 214
```

RESULT 15

30989

serine/threonine protein kinase NIK - mouse
N;Alternate names: Nck interacting kinase

Species: *Mus musculus* (house mouse)

```
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
```

Accession: T3098

R; Su, Y. C.; Han, J.; Xu, S.; Cobb, M.; Skolnik, E. Y.
EMBO J. 16, 1279-1290, 1997

